DOCUMENT RESUME

ED 243 905 TM 830 843

TITLE	Archiving Methodology. Volume II: Contractor's Guide.
INSTITUTION	Leinwand (C.M.) Associates, Inc., Newton, Mass.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	31 Mar 80
NOTE	29p.; For related documents, see TM 830 842-845.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Archives; *Databases; Data Collection; *Delivery Systems; Documentation; Guidelines; Information Utilization
IDENTIFIERS	*Contractors; Database Producers; *Data Organization; Secondary Analysis

ABSTRACT

This section of the "Archiving Methodology" is concerned with the role of data collection contractors in the archiving process. Their role is discussed from three perspectives: (1) the nature of the relationship of the data collection contractor, the project officer, and the archivist; (2) the considerations contractors must make when preparing data for archiving; and (3) the mechanisms facilitating the transfer of information from the contractor to the archivist. This document discusses an overview of contractor requirements (project-and file-level documentation) and specific contractor requirements (Cross-Referenced Information Form and data collection process considerations). (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED243905

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ARCHIVING METHODOLOGY

VOLUME II: CONTRACTOR'S GUIDE

Submitted to

National Institute of Education

by

C.M. Leinwand Associates, Inc.
March 31, 1980

TM 830 813

CONTENTS

I. AN OVERVIEW OF CONTRACTOR REQUIREMENTS 4

A. PROJECT-LEVEL DOCUMENTATION 4

B. FILE-LEVEL DOCUMENTATION 5

II. SPECIFIC CONTRACTOR REQUIREMENTS 6

A. Cross-Referenced Information Form (CRIF) 6

1. CRIF: Single-Study Projects 8

2. CRIF: Projects Consisting of Multiple Substudies 8

3. CRIF: Individual Study Files 11

B. Data Collection Process Considerations 11

1. Initial Study Considerations 14

2. Ongoing Collection Considerations 17

3. Post-Collection Considerations 20

FIGURES

1. Cross-Reference Information Form: Single-Study Projects 9

2. Cross-Reference Information Form: Projects Consisting of Multiple Substudies 10

3. Cross-Reference Information Form: Individual Study Files 12

VOLUME II

GUIDE FOR THE DATA COLLECTION CONTRACTOR

This section of the Archiving Methodology is concerned with the role of data collection contractors in the archiving process. Their role is discussed from three perspectives:

- the nature of the relationship of the data collection contractor, the project officer, and the archivist;
- the considerations contractors must make when preparing data for archiving;
- the mechanisms facilitating the transfer of information from the contractor to the archivist.

As Volume I indicated, the project officer decides what should be archived, how much effort should be devoted to archiving it, and how the information in the final product (archive) is to be disseminated. The contractor and archivist perform their work according to these decisions; the final product--the data archive--conforms to the guidelines the project officer has established.

The archivist depends on the contractor to provide study information sufficient to document the study data.

Preparing an archive depends on a reciprocal relationship between the data collection contractor and the archivist. Like any good working relationship, this relationship demands communication between the parties involved. The archivist stipulates the kinds of products required from the contractor for the creation of usable and appropriate documentation. In turn, the products the contractor delivers must be sufficiently detailed and comprehensive to describe the data collection effort adequately. To facilitate this cooperative effort, we will discuss the materials the archivist generally requires from the contractor. These include copies of all reports and tapes generated by the study.

Two issues are fundamental to understanding contractor requirements. First, the materials needed are those expected from, and generated by, any sound research and reporting activity. Second, the recording of the necessary information takes place throughout the course of the study.

The archivist does not ask the contractor for materials and information peculiar to the archiving endeavor or superfluous to the research process. For example, in requesting a detailed sampling plan including modifications made and proportions of the target population obtained, the archivist is asking for information that vitally affects the data. Normally, the contractor records this information as the study is being planned and as data collection proceeds, and later incorporates it in the study reports. The contractor provides this information for the same reason the archivist desires it: the data are affected by sampling design and obtained population. Likewise, readers of the study reports and users of a data archive seek this information for the same purpose: to understand what the data represent.

In reality, studies do not always meet the ideal of sound research and reporting practices. To the extent that a study departs from this ideal, the archivist must request "additional" products from the contractor. However, in such cases, it is clear the archivist is helping the contractor more completely report the research conducted. As a result, the archiving process will affect the study beneficially. If contractors know the data generated by their studies are to be archived, they will be encouraged to maintain logs and records accounting for study design modifications and data anomalies. Cooperation between the contractor and the archivist results in not only a useful repository for data, but also an improved research effort.

Before addressing the requirements that archiving places on the contractor, we will reiterate and emphasize the major points of the discussion above to orient the contractor appropriately to the archiving process.

- The archivist depends on the contractor for all information affecting the data.
- Cooperation from the contractor in delivering products to the archivist is essential to the creation of a thoroughly documented data archive.
- Whenever in doubt, the contractor should provide the archivist with more rather than less study information.
- The information the contractor delivers to the archivist is consistent with the requirements of good research conduct and reporting.
- Throughout the study, contractors should carefully record all deviations from the proposed methodology (i.e., sampling plans, obtained populations, instrumentation, coding practices, etc.).
- In addition to the formation of a data archive, the joint efforts of contractor and archivist will help to improve the research process by encouraging and making available more complete and accurate accounts of the study.

I. AN OVERVIEW OF CONTRACTOR REQUIREMENTS

The information the archivist requires from the contractor is determined by a set of standards which have been developed for the preparation of a data archive. These standards are described in detail in Volumes III and IV of the Archiving Methodology. Briefly, there are three levels of archive documentation: project, file, and programmer. The contractor's role in data archiving is focused on the project and file levels.

A. PROJECT-LEVEL DOCUMENTATION

Project-level documentation is the most global level; it is usually an archive user's introduction to a specific data archive. The goal of project-level documentation is to acquaint the reader, in a single volume, with the study that generated the data in the archive and to provide the reader with such information as the study's purpose, its design, its sampling strategy, and its key results. It also describes all the substudies undertaken as part of a project. Reading the project-level documentation will help the potential researcher to decide whether the study may fit his/her needs.

To produce an adequate project-level document, the archivist requires the proposal and all overview documents generated by the study. Since documentation at this level is relatively general, very specific details of study modications are not required. In fact, a project-level document would fail to fulfill its primary purpose--namely, to give the potential researcher a general understanding of the study--if all specific data were included in it. This information is more appropriately presented in the descriptions of the individual file(s) generated by the study.

B. FILE-LEVEL DOCUMENTATION

If the project-level document suggests that the study fits a researcher's needs, s/he turns to the file-level documentation for more specific detail. A file-level document is written for each separate data file or each group of identically structured files. These documents are the analyst's guides to using the data files. Usually, data generated by a single instrument comprise a single file. However, in the case of an instrument administered three times, the data from each administration may reside on a separate physical file. If each administration produced the same data, a single file-level document is sufficient. Separate file-level documents are required when instruments differ substantially from one administration to another.

To produce file-level documents, the archivist requires precise information about each file. Often, a study consists of multiple data collection efforts. For example, the Elementary Education Voucher Demonstration (EEVD) surveyed parents, teachers, principals, etc. Each of these data collection efforts relates to the overall purpose of the EEVD in a slightly different way and contributes to the answers of the overall research questions. The archivist must understand these purposes and their relationship to the overall project purpose in order to explain the data file adequately. Similar detail is required about all the remaining components of the file-level standard (unit of observation, scale, time information, sampling and target population, data collection method, data anomalies, etc.).

The following section will enlarge upon these general contractor requirements and provide a formal mechanism to aid the contractor in delivering them to the archivist.

II. SPECIFIC CONTRACTOR REQUIREMENTS

The process of data collection can be viewed chronologically; that is, in terms of the considerations a contractor must make about a study while it is being planned, initiated, and implemented and in retrospect when it is completed. When a study's data are intended for a data archive, the contractor must be aware that the archivist is also interested in what is occurring during each of these data collection phases. Again, we emphasize that these considerations normally occur in the course of any research study, regardless of whether the data is to be archived. The only "burden" archiving places on the contractor is that of communicating these decisions to the archivist.

A. CROSS-REFERENCED INFORMATION FORM (CRIF)

Although archiving requires little or no additional work for the contractor, other than the communication of study information to the archivist, it is necessary for the contractor to use a mechanism to facilitate such communication. We propose a "Cross-Referenced Information Form" (CRIF) be used as this mechanism because it meets several criteria.

- It has been designed to provide the archivist with all the information needed to reconstruct the study on both the project and file levels.
- Despite its comprehensiveness, CRIF is a simple form; completing it will not demand a great expenditure of the contractor's time and energy.
- It is a clear, concise, easy-to-use form, which lists and asks the contractor for information on a set of elements corresponding to all phases of the study. The contractor responds by using CRIF to refer the archivist to certain pages of a study publication.

The assumptions underlying CRIF are that all information required to archive data can be found in the literature the study generated and that the contractor is not being asked for anything that does not already exist. The information required to complete the CRIF may come from many sources: the original study proposal, quarterly or interim reports, OMB clearance statements, final reports, appendices, and other study documents. It is important to note that the development of the documentation is the archivist's role; the contractor is being asked to identify the location of the information needed for the documentation.

At the beginning of the study, the archivist supplies the contractor with copies of different versions of CRIF. Consequently, the information the archivist requires will be known to the contractor far in advance of its due date. As the study progresses, the contractor can gradually complete the CRIF. An advantage of this approach is that it can help the contractor determine if all the necessary information exists in the study's reports and documents. Used in this way, CRIF is a useful checklist which not only supplies necessary information to the archivist but also reflects the essential components of a research study. If the contractor encounters great difficulty in completing the CRIF, this can be interpreted as an indication to the contractor that certain areas of the study have not been adequately discussed in study documents, or, more seriously, that certain essential research considerations have not been made.

Several versions of CRIF appear in the next section. The reader will readily observe that the elements comprising CRIF are essential components of all research studies. CRIF does not disrupt the nature of the process, but rather reflects it faithfully.

On the next few pages, we present several versions of CRIF. Our reasons for providing these versions relate to the kinds of information the archivist needs from the contractor to generate project- and file-level documents. The contractor completes both project-level and file-level CRIFs, indicating where relevant information may be found. The file-level CRIF differs from the project-level CRIF because more specific detail is needed for file-level documents. When the data archive is established, researchers will use the file-level documents in their work with separate files. To work effectively with a file, these researchers will need precise information about its contents, source, and format.

In addition, certain projects are overarching research endeavors, composed of multiple substudies. Such large-scale studies require a project-level CRIF appropriately organized to present the location of information about the overall project and its component substudies. For this reason, two versions of the project-level CRIF are given below.

1. CRIF: Single-Study Projects

The single-study project CRIF shown on the next page (Figure 1) is a relatively straightforward listing of study components. When the contractor has cited the locations of information about each component, the archivist has the information required to generate a project-level document. The archivist will decide how much information cited is necessary to explain the project adequately.

2. CRIF: Projects Consisting of Multiple Substudies

The version of CRIF shown in Figure 2 is used when the data being archived represent large scale research project involving more than one related

FIGURE 1

Cross Referenced Information Form: Single Study Projects

Study Title: _____

	Location	
	Document	Page #
A. Rationale for Data Collection Effort		
1. Purpose(s) of study		
2. Major research questions		
3. Historical/theoretical background		
4. Study hypotheses		
B. Study Design		
Major study variables		
C. Method of Data Collection		
D. Data Analysis		
E. Major Findings		

FIGURE 2

Cross-Referenced Information Form: Projects Consisting of Multiple Substudies

Study Title: _____

	Location	
	Document	Page #

A. Rationale for Project

1. Purpose(s) of project
2. Major research questions
3. Historical/theoretical background
4. Major hypotheses

Substudy I....K

Substudy Title: _____

A. Rationale for Substudy

1. Relationship to overall project
2. Purpose(s) of substudy

B. Substudy Design

C. Method of Data Collection

D. Data Analysis

E. Major Findings

substudy. By completing this form, the contractor tells the archivist where in the project literature to find the information necessary to produce a project-level document. Such a document acquaints the reader with general information about the project and the separate data collection efforts (substudies) it encompasses. Items lettered I- K and A - E are completed for each substudy within the project.

3. CRIF: Individual Study Files

The future archive user requires more precise information on the file-level because it is this documentation which will enable the user to actually work with the data. The CRIF for individual study files (Figure 3) reflects this need for greater specificity.

In this version of CRIF, research components are organized under the three headings: initial, ongoing, and post-data collection considerations. These chronological distinctions are made to elicit facts about any deviations from the original data collection plans. As with the previous versions of the CRIF, the contractor provides only the location of this information; it is up to the archivist to interpret the information cited and decide what is needed to generate a complete file-level document.

B. DATA COLLECTION PROCESS CONSIDERATIONS

This Methodology maintains the viewpoint that every decision made in data collection in some way affects the data themselves. The purpose of this section is not to instruct the contractor in the conduct of research, but rather to call the contractor's attention to certain issues critical to archiving which are often neglected or treated haphazardly in study reports.

FIGURE 3

Cross-Referenced Information Form: Individual Study Files

File I.....K

Study Title: _____

File Title: _____

	Location	
	Document	Page #
I. Initial Considerations		
A. Rationale for Data Collection		
1. Purpose(s)		
2. Relationship to other study files and to study as a whole		
B. Unit of Analysis		
C. Design of Instrument		
1. Validity		
2. Reliability		
D. Sample Charactertistics		
1. Size/characteristics of universe		
2. Target population		
E. Sampling Strategy		
F. Time Information		
1. When?		
2. How often?		

	Location	
	Document	Page #

II. Ongoing Study Considerations

A. Time Information Modifications

B. Sampling Strategy Modifications

C. Interviews/Unusual Occurrences

D. Instrument Modifications

III. Post-Collection Considerations

A. Scale

B. Obtained Response

C. Coding Decisions

D. Data Editing/Cleaning

E. Weighting of Data

The following is a brief description of the important elements in CRIF and an explanation of how these elements will assist future users of the archive. The elements are grouped chronologically into initial study considerations, ongoing collection considerations, and post-collection considerations.

1. Initial Study Considerations

Information about initial study considerations are likely to be found in the enabling legislation, the request for proposals (RFPs), the proposal, the contract, and early status reports on the study, as well as in more "final" documents. They include

- rationale for data collection;
- study design;
- instrument design;
- validity;
- reliability;
- sample characteristics, size, target population;
- sampling strategy;
- time information.

a. Rationale for Data Collection

Every data collection effort is designed to meet some expressed purpose(s). The reason(s) for obtaining the data should be stated clearly, specifying the historical and theoretical background of the issue under study. The specific theoretical stance which guided the study and the hypotheses it tested are major considerations in constructing a comprehensive account of why the data were collected.

b. Study Design

The study design grows out of the specific theory being tested: it is a blueprint diagramming the way study variables realte. These relationships form the theoretical model being tested. The study design identifies the variables of interest or the independent/dependent variables when appropriate. The operational definitions of study variables and the unit of analysis can also be included here.

Future archive users must be informed of the design of the study from which the data originated, since such information explains the meaning of the data in the context of the previous study. This information will assist the secondary analyst in determining areas of theoretical weakness.

c. Instrument Design

Usually, survey data are collected with some type of data collection instrument. Non-survey data are obtained using a form or the output of an administrative system. It is important to archive users to know how data were generated and, in the case of multiple administrations, to know what changes, if any, were made to these instruments over time.

d. Validity

A valid instrument is one which, in fact, measures what it purports to measure. Valid instruments are essential tools of sound research. There are several types of validity: face, content, criterion-referenced, and construct. Surveys are most likely to strive only for face and content validity. The former type of validity refers to the appearance of the instrument, while the latter concerns the adequate sampling of a clearly specified subject area. Criterion-referenced validity is established when

an instrument is shown to correlate well with other criteria of a trait, attitude, skill, etc. Finally, construct validity is the testing of the underlying theory of an instrument. To the extent that the theory is upheld, an instrument is content valid.

Information about the validation procedures used and their outcomes enable the future researcher determine if further testing is needed to establish an instrument's validity. In cases where validation was not performed, future users may perform their own.

e. Reliability

The reliability of an instrument refers to its accuracy, precision, and replicability as a measurement tool. As with validation procedures, if reliability indices were constructed, future users will be interested in their results and the techniques used. Where several waves of data are being archived, the precision of the measuring instrument must be taken into account when comparing the results of one administration to those of another.

f. Sample Characteristics

Size characteristics of the universe. The universe is the population from which a sample is drawn. Secondary analysts need this type of documentation to help them generalize their results to the correct population.

Target population. This component describes the type of sample drawn, its important characteristics, and the reasons for choosing it. The target group represents the "goal" population toward which data collection efforts strive. A thorough description of the target population acquaints the archive user with this group.

g. Sampling Strategy

A sampling strategy is the specific selection criteria and methodology used to draw a sample appropriate to a study's objectives. The description of this strategy gives future researchers an opportunity to evaluate the sample's suitability for answering the research questions the study poses.

h. Time Information

As part of the proposal and initial study considerations, the contractor prepares a data collection schedule. It specifies when the data will be collected (month, season, year) and how often (once, many times, or continually). The actual time of data collection may explain some seasonal effects; the frequency of data collection can suggest possible bias as a result of attrition and/or the possibility of correlated results owing to repeated measures.

Data usually refer to a specific time period in either the present or past. This specific time period, the "data time frame," is not always obvious, especially if attitude is being measured. The data time frame should be clearly stated to users of the data. Often, a respondent is asked to express a presently-held attitude about some past event, phenomena, object or person, for example, "How satisified are you with the special services which your child received last year at school?" In such cases, the time frame is considered present.

2. Ongoing Collection Considerations

Ongoing collection considerations might be called "modifications and problems," since most of these considerations become important when, for practical reasons, it is not feasible or possible to adhere to all of the initial study considerations. Included in ongoing considerations are

- time information modifications;
- sampling strategy modifications;
- interviews and other unusual occurrences in data collection;
- instrument modifications.

a. Time Information Modifications

We discussed this factor as an initial consideration; we discuss it again here because the data collection factors of "when?" and "how often?" are sometimes a concern throughout the collection process. Time schedules are most likely to deviate because segments of the target population are difficult to reach, interviews can take longer to administer than planned, mailed questionnaires may be returned later than expected, and many other interruptions can occur. Significant deviations from the proposed data collection schedule can profoundly affect the resultant data, especially deviations involving a period of months or failure to conduct a proposed wave of data collection.

b. Sampling Strategy Modifications

Studies can initially propose to draw a sample of a certain size and type and later modify the earlier strategy. There may be many reasons for modification, among them, practical factors, such as inability to contact desired segments of the population, or recognition that the original sampling plan was overly idealistic or excessively expensive. An alternative sampling strategy may be adopted because unanticipated substantive issues arise which require selecting a new or larger set of respondents. The rationale for employing a new or modified sampling strategy and a detailed description of the alternative strategy should be included, since without this information,

the archivist is unable to adequately convey the rationale underlying the data in the archived file.

c. Interviews and Other Unusual Occurrences in Data Collection

Interviewing is a common technique for collecting data. Unlike the mailed questionnaire which is a relatively controlled stimulus for response, the interview situation may encourage variation, primarily because of interviewer interference or bias. Interviewer bias can contribute significantly to response error and is difficult to control even with the most ideal training conditions. The interpersonal reaction between interviewer and respondent is another factor which can influence the nature of responses. A host of other factors can affect the responses obtained, among them, length of interview; convenience of interview appointment for respondent; interviewer training and experience; language competency of respondent; cultural characteristics of interviewer and respondent; relevancy of interview to respondent; and clarity of the questions asked.

When conducting a study which uses the interview for data collection, the contractor must be aware of all of the above factors and carefully note any factors which seem to influence the data collection process. A useful technique some studies employ is to ask interviewers to report to a supervisor periodically to discuss problems and any peculiarities encountered in the set of interviews completed. Since such problems and peculiarities are important to the archivist, contractors can instruct the supervisor to maintain records of them. Here, as with other deliverables, the contractor can rely on the advice that it is better to provide the archivist with "more rather than less" study information. Although much of this information may not ultimately be included in the archive, it will enable the archivist to make more informed decisions about what is appropriate explanatory material.

Any other unusual events or problems which occur during the data collection process are of interest and should be recorded here. Internal events, such as staff change or funding problems, may have a bearing on data collection. "External" occurences, such as strikes, elections, community activism, organizational changes, natural or economic disasters, and other events in the environment may be relevant to the issues being investigated by the study.

d. Instrument Modification

When a survey is administered in waves, it may become evident that items in the data collection instrument are ambiguous or troublesome to respondents. In addition, issues are subject to change, especially in longitudinal studies and in large-scale environmental studies. For these reasons, items may be deleted, revised, or added. Since the instruments are the tools which generate the data, it is essential that all instrument modifications be clearly described. An explanation of the rationale for changes in the study instruments can help explain why the data may differ from administration to administration.

3. Post-Collection Considerations

These are the considerations made after the data have been collected. They relate to how the data are to be analyzed to answer the research questions. Of course, all of these considerations were also made in the beginning of the study. However, the reality of the collected data may not be consistent with original expectations for it. Hence, many of the following issues need reconsideration; others arise for the first time:

- scale;
- obtained response;
- coding decisions;

- data editing and cleaning;
- weighting of data;
- analysis processes;
- proposed techniques;
- index construction;
- actual statistical techniques;
- data anomalies and limitations;
- major findings;
- file processing considerations.

a. Scale

Scale is the size of the data. There are two components of size: 1) the number of cases, that is, the actual number of records on the file; and 2) the number of variables collected, that is, the number of data fields each record contains. Scale information is important to the archivist, since secondary analysts and other users of the archive will have to determine whether a file is an appropriate size for their needs.

b. Obtained Response

The description of the characteristics and size of the data file can differ substantially from the intended target population. The obtained population represents the true contents of the data file. The differences between the obtained and target populations most commonly occur because of problems encountered in data collection. Some typical problems are that respondents were not at home, they refused to participate, or questionnaires were mailed to their old addresses. Respondents can misinterpret questions which necessitate the removal of their responses from the sample.

All significant discrepancies between the obtained and target populations should be reported to the archivist. This information will help future researchers understand any bias to data due to nonresponse.

c. Coding Decisions

Coding practices and decisions about how to treat open-ended responses, in particular, vitally affects the meaning of the data. Contractors should complete the appropriate sections of the CRIF at the time the coding decisions are made. If this is done, the archivist will not need to rely on the fading memories of individuals involved in the study and will not have to engage in an ex post facto accounting of study decisions; instead, the archivist will receive complete "on the spot" reports. It cannot be overly emphasized that contractors should work on the CRIF as the study progresses. Experience in archiving has shown that researchers, busy with present projects, have great difficulty remembering the details of studies in which they were previously involved.

d. Data Editing and Cleaning

Data editing and cleaning consistency checks performed on data include syntactic and semantic checks and the imputation of errors or missing values. Syntactic checks concern the form and charactertistics of an individual data field and insure that each field conforms to individually defined characteristics. For example, if a SEX field were specified as alphabetical, with acceptable values of "M," "F," or "blank," a syntactic check would be used to ascertain that no other values appeared in the field. The semantic check investigates relationships between two or more variables and insures that data within a record is consistent and reasonable. Thus, an example

of a semantic check would be assuring that a student with a GRADE of 1 must have an AGE between 4 and 8.

After errors or missing values have been detected, the contractor must decide how they should be treated. The contractor can choose to leave the decision up to the user of the data or to employ some sort of correction procedure, such as clerical review, correspondence, re-observation, or automatic data correction. If new values or imputations have been used to replace old or missing values in the data, the contractor should cite in the CRIF the location of this information in the study literature. Future archive users require this information in order to gauge the appropriateness and the magnitude of the imputations made in the data.

Information on the checks performed on the data and their results will inform the archive user that the range of values is legitimate or signal the existence of problems within the data.

e. Weighting of Data

During the course of the study, the researcher may wish to estimate population parameters or establish relationships requiring population parameters. Since samples often contain one or more substrata of the population that have been over- or under-sampled, weights can be assigned to cases, that is, some individuals may be weighed more or less heavily to approximate the characteristics of the population.

A particularly important concern with weighting in secondary analysis is the population for whom this weighting system was devised. The contractor must clearly indicate whether the weights are intended for national, state or local estimates.

Future users of an archive containing data designed to obtain national estimates are spared time and energy if they are informed of the weighting

schemes employed in previous studies. However, as population characteristics change over time, future users should cautiously approach previously used weighting schemes, particularly if the schemes were developed several years prior.

f. Analysis Process Considerations

As with the considerations made for data collection, the archivist is interested not only in proposed data analysis techniques, but also deviations from the initial analysis plans. Of particular importance to the archivist are anomalies found in the data as a result of the analytic process. The actual results of data analysis are not of primary interest to the archivist unless these results point to the existence of irregularities within the data.

The following is a description of the CRIF elements which apply to data analysis.

Proposed Techniques. Data analysis, in its most general usage, involves numerical description, estimation, and inference with regard to a set of data. Numerous statistical techniques can be performed on a data set. They range in complexity from simple summary statistics, measure of central tendency and dispersion, to more complex, multiple regression techniques. Regardless of the technique employed, the ultimate purpose of data analysis is to answer the questions posed by the research study.

Knowledge of the types of statistical analyses planned for a data set will enable future researchers to decide if additional kinds of analyses could be used to enrich and extend the work of the original contractor and to provide further insights into the data. Scrutiny of statistical techniques

can uncover possible misapplications of techniques and serve as a check on the results of the original study.

Extensive descriptions of statistical methods used are not required, since such information is of relatively limited value to future users of the archive. If users require detail about the statistical methods employed in the study, this information can be obtained from the final study report.

Index Construction. In addition to data analyses, the original researcher team may build indices which are measures resulting from the combination of the values of several items or variables. An index is usually a more efficient gauge of a construct than is a single item or variable.

These data manipulations are important in a study, since they explain the theoretical structure of the variables. As with data analysis, the indices employed should be described briefly; detailed accounts of these data manipulations are not required.

Actual Statistical Techniques Performed. During the course of a study, additional analyses may be undertaken. This decision may be made because the results obtained using the initially proposed techniques suggest that more analyses be performed or because the nature of the original expectations about the data have changed. To provide an accurate account of the statistics used in the study, information about the deviations from the previously planned techniques are used.

Data Anomalies and Limitations. Problems and anomalies arising from the data collection effort can include incomplete samples, anomalies in the instrument or its accompanying instructions, and adjustments made to the data after collection. If problems, anomalies, and the strategies for

treating them are not clearly explained, the result may be improper interpretation of the data. Problems and anomalies can be uncovered during data collection, editing, and cleaning or as a result of the data analysis process. To properly document the data for future archive users, the archivist requires this information.

g. Major Findings

The findings of interest to the archivist are those indicating whether or not the study supported the hypothesis. These findings are recounted in the project-level documentation, which gives the potential user an overview of the project. The significance of these findings and their relationship to theory are not required by the archivist. Future researchers who desire this information can consult the final report cited in the bibliography accompanying each study.